Pulsed versus Continuous Tones for Evaluating the Loudness of Tinnitus

James A. Henry*
Mary B. Meikle*

Abstract

Loudness balance techniques are commonly employed to match the loudness of tinnitus using either pulsed or continuous tones; however, it is not known whether the tone duration affects the observed loudness matches. In this study, hearing thresholds and tinnitus loudness matches were measured in 26 subjects with chronic tinnitus using both pulsed and continuous tones. Subjects' thresholds and loudness matches were determined at 11 frequencies between 0.5 and 10 kHz. No significant differences were found between pulsed versus continuous measures, either for thresholds or for loudness matches. There were, however, nine subjects (34.5% of the group) who showed relatively large differences (>10 dB) at one or more test frequencies. These “outlier” values did not show systematic trends; some were positive, some negative. In conclusion, studies employing group data appear to be comparable if group sizes are sufficiently large (e.g., ≥ 25 subjects). Studies employing smaller numbers of subjects may, however, be vulnerable to potential positive or negative biases introduced by one or more outliers.

Key Words: Continuous, high-frequency hearing, intermittent, loudness, pulsed, tinnitus

Abbreviations: ANOVA = analysis of variance

Subjective tinnitus is well known to accompany many forms of auditory dysfunction. It is also known that severe tinnitus can interfere with an individual’s ability to function in one or more aspects of daily life (work, leisure time or social activities, sleep, etc.) (Evered and Lawrenson, 1981; McFadden and Wightman, 1983; Vernon and Meikle, 1985). Recent information indicates that at least 10 million Americans suffer severely from tinnitus (Brown, 1990); thus, it is an important health problem.

Historically, work in the area of tinnitus was sporadic until the 1970s, when several new techniques for treatment were developed (Vernon, 1975, 1977; House et al, 1977; Melding et al, 1978; Hazell, 1987). Since that time, there has been a dramatic increase in the number of studies devoted to tinnitus and a worldwide effort to improve both measurement and treatment (Shulman, 1981; Hazell, 1981; Feldman, 1987; Kita-hara, 1988; Shulman et al, 1991; Aran and Dauman, 1992). Evaluation of treatment efficacy frequently involves techniques for quantifying the perceptual attributes of tinnitus (Vernon et al, 1977; Hazell, 1981; Vernon and Meikle, 1981, 1988; Tyler and Conrad-Armes, 1983; Vernon, 1987). In general, these techniques involve measurement of such attributes as the pitch, loudness, and maskability of tinnitus. However, specific measurement protocols have differed between investigations.

One important method for evaluating the loudness of tinnitus is based on Fowler’s (1942) method, whereby an external tone is adjusted to match the loudness of an individual’s tinnitus. Such a method does not actually measure the subjective or perceived loudness of tinnitus but rather measures the level of an equivalent external sound (usually a pure tone). The advantages of this method are that it is more objective than methods involving subjective magnitude estimation, and the results have been reported to be highly reliable (Bailey, 1979; Vernon et al, 1980). However, a variety of different techniques...
have been used to obtain loudness matches for tinnitus (Goodwin and Johnson, 1980; Penner, 1983; Tyler and Conrad-Armes, 1983; Coles et al, 1984; Jakes et al, 1986; Vernon, 1987; Matsuhira et al, 1992; Mitchell et al, 1993), and some investigators have reported poor reliability of responses. Differences between studies have included choice of test frequencies, stimulus duration, subject selection criteria, choice of test ear, and the nature of the psychophysical task employed. These differences make it difficult to compare results from different testing sites.

Fowler (1937) introduced the loudness matching technique using continuous tones, and most studies have followed suit. However, some investigators have used pulsed tones. Because the two techniques have not been compared under controlled conditions, it is not clear whether tinnitus loudness measures using pulsed tones and measures using continuous tones are equivalent.

Two previous investigations have been conducted to compare pulsed versus continuous tones for threshold evaluations in individuals with tinnitus (Hochberg and Waltzman, 1972; Mineau and Schlauch, 1997). In general, pulsed tone thresholds were lower than continuous tone thresholds, although the differences were not statistically significant in either study. To our knowledge, there is no previous study comparing the use of pulsed versus continuous tones for obtaining tinnitus loudness matches.

Therefore, the major purpose of the present study was to compare tinnitus loudness matches obtained with pulsed tones to those obtained with continuous tones, in order to determine whether there are systematic differences between the two methods. Additional information was also obtained concerning pulsed versus continuous thresholds in the same group of tinnitus patients.

METHOD

Subject Selection

Subjects for this study were enlisted from the group of patients who had previously been evaluated at the Oregon Tinnitus Clinic. Patients were selected to meet the following criteria: tonal tinnitus that is not intermittent, is louder on one side of the head than the other (to help avoid confusion when external matching tones were presented contralaterally), and with pitch matched to 3 kHz or 4 kHz. (This last requirement was included to increase the probability that tinnitus loudness matches could be obtained at frequencies both above and below the individual's tinnitus frequency.)

Equipment

All psychoacoustic evaluation was done in a double-walled, sound-attenuated chamber using a Virtual Corp. (Portland, OR) Model 320 audiometer under computer control. In addition to the conventional-frequency range (0.25–8 kHz), the 320 provides valid and reliable thresholds in the high-frequency range (9–20 kHz) (Fausti et al, 1990, 1993). The computer software controlling the audiometer normally allows evaluation of hearing thresholds in 5-dB steps; however, for this research, the software was modified by the manufacturer to enable threshold and loudness evaluation in 1-dB steps at 11 test frequencies (0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, 9, and 10 kHz).

Test stimuli were presented using TDH-50P earphones in MX-41/AR cushions. Although these earphones are standard components of the Model 320 system for conventional frequencies, they can also be calibrated for use with the 320 at 9 and 10 kHz; however, their sound output is reduced at those frequencies. For subjects with severe hearing loss who required greater output at 9 to 10 kHz, Koss Pro4/X Plus earphones (provided with the Model 320 for testing in the range above 8 kHz) were used for those two frequencies. Calibration with the TDH-50P earphones was done in dB HL (ANSI, 1989) for frequencies in the range 0.5 to 8 kHz and in dB SPL at 9 and 10 kHz. The Koss earphones were calibrated in dB SPL using the flat-plate coupler method described by Fausti et al (1979, 1990).

Characteristics of the pulsed tones were within the American National Standards Institute's Specification for Audiometers (ANSI, 1989), with on-times of approximately 185 msec and rise–fall times of approximately 37 msec. The repetition rate was 2.5 per second. Continuous tones used as test stimuli were sustained until the subject responded or, if no response was obtained, for a maximum period of approximately 2 seconds.

Testing Protocol

The experimental protocol involved four sets of measurements for each subject, consist-
Pulsed versus Continuous Tinnitus Evaluation/Henry and Meikle

ing of (1) hearing thresholds and (2) tinnitus loudness matches, each obtained under two conditions (pulsed vs continuous stimuli). To control for possible order effects, half of the subjects (selected randomly) received threshold and loudness testing in the pulsed mode first, followed by the continuous mode; the order was reversed in the other half of the subjects.

The procedures used to measure tinnitus loudness were adapted from those used at the Oregon Tinnitus Clinic (Vernon, 1987). During testing, the examiner sat inside the Industrial Acoustics Corporation chamber, facing the subject, as this arrangement facilitates the necessary interaction between the examiner and the subject (J. Vernon, personal communication). To eliminate interference created by noise from the computer, the external hard-disk drive was located outside of the test chamber. The subject was not permitted to see the monitor screen of the computer, and care was taken to prevent visual or auditory cues that might have biased the subject's responding during either threshold or loudness testing.

Prior to testing, subjects were asked to identify the ear with the predominant tinnitus, hereafter referred to as the “tinnitus ear.” Test stimuli were presented to the contralateral ear (the “test ear”). For each stimulus mode, testing began at 500 Hz and continued through 10 kHz in order of ascending frequency. At each frequency, hearing threshold was bracketed within a 5-dB interval using the Hughson-Westlake ascending method (Carhart and Jerger, 1959). In order to achieve precise 1-dB resolution, the stimulus level was then adjusted in 1-dB steps within that interval. Final 1-dB threshold responses were verified by repeating this down-5-up-1 process.

To match the loudness of tinnitus, the sound level was increased in 5-dB steps above threshold until the subject reported that the test tone was nearly equal in loudness to the tinnitus. The signal was then increased in 1-dB ascending steps until the subject reported that the external sound was equal in loudness to the tinnitus. The final level of the external sound was recorded as the tinnitus loudness match (in dB) for that frequency.

Data Analysis

In the conventional-frequency range (0.25–8 kHz), the Virtual audiometer records subject responses in dB HL. For frequencies above 8 kHz, responses are measured in dB SPL. In order to compare measurements at the various test frequencies using a common metric, the data obtained at 0.25 to 8 kHz were converted to dB SPL using the published correction at each frequency (ANSI, 1989). During statistical analysis, the loudness matches were evaluated both in terms of dB SPL and in dB SL above threshold. Statistical analysis of the data was done using SPSS/PC+ (v. 6.2, SPSS Inc.) (Norusis, 1986), SuperANOVA™ (v. 1.11, Abacus Concepts), and Statview II™ (v. 5.0, Abacus Concepts).

RESULTS

Subject Information

A total of 26 subjects provided data for this study. The sample consisted of 21 men and five women, ranging in age from 34 to 84 years, with a mean age of 58.7 years. As is typical for tinnitus clinic patients, individuals in this group tended to have substantial hearing loss that was greatest at frequencies above about 2 kHz.

Information concerning the most probable cause(s) of the subjects’ tinnitus is summarized as follows: (1) 11 subjects reported sudden-onset tinnitus. Of those 11, 7 reported that their tinnitus was caused by brief intense noise (5 with and 2 without additional complicating factors such as occupational noise exposure), and the other 4 reported non-noise-related causal factors (whiplash, illness, head injury, ear surgery); (2) the remaining 15 subjects reported gradual-onset tinnitus. Twelve of those subjects reported long-term noise exposure (occupational, military, and/or recreational) and also had audiograms consistent with noise-induced sensorineural hearing loss. The remaining three patients had no noise exposure or other evidence indicating any specific causal factor(s) for the tinnitus.

Testing for Order Effect

Two-factor analysis of variance (ANOVA) (presentation order × stimulus frequency) was used to determine whether there were significant response differences between the group of subjects who received pulsed stimuli first versus those who received continuous stimuli first (N = 13 in both groups). No significant differences were found in main effects or interactions either for thresholds or for loudness matches (p > .20); therefore, data for all 26 subjects were combined for further analysis.
Pulsed versus Continuous
Hearing Thresholds

Mean Thresholds

Thresholds obtained with the two stimulation modes are summarized in Table 1, where it can be seen that the absolute difference between mean thresholds does not exceed 1.3 dB at any of the 11 test frequencies. Mean thresholds for the two stimulus modes were compared using a 2 x 9 ANOVA (stimulus mode x frequency) with repeated measures on both factors. As would be expected in any population with significant high-frequency hearing impairment, there were significant differences between the mean thresholds at different frequencies. However, the mean thresholds for pulsed versus continuous tones did not differ significantly (F [1, 25] = 0.001, p = .978).

Individual Data

Individual difference scores were computed at each test frequency by subtracting each subject’s continuous threshold from the pulsed threshold. Figure 1 displays the difference scores for the entire group of subjects in the form of 11 frequency histograms stacked vertically with the histogram for the 500-Hz difference scores at the top and the histogram for the 10,000-Hz difference scores at the bottom. The data displayed in this figure make it possible to evaluate individual differences in regard to thresholds obtained with pulsed versus continuous tones. In general, most difference scores were in the range ±5 dB, except at 750 Hz, where most scores were in the range ±7 dB. A few “outliers” falling outside the range ±10 dB are visible at 1500 Hz and also at 4, 6, 8, and 10 kHz.

Pearson’s product-moment correlations were computed for the pairs of values represented by each subject’s corresponding pulsed and continuous thresholds, in order to quantify the extent of linear relationship between responses obtained using the two stimulus modalities. The high degree of linear correlation between the pulsed and continuous thresholds is clearly evident from Figure 2, which shows scatter plots for four representative frequencies (0.5, 2, 4, and 8 kHz). Most values can be seen to lie within a few decibels of the regression line. These conclusions can be evaluated somewhat more quantitatively if we look at Table 2, which shows the correlation coefficients at each frequency and also provides the standard errors of estimate for pulsed thresholds, using continuous thresholds as predictors (Downie and Heath, 1965). For the present sample of subjects, the standard errors of estimate are close to 3 dB for the frequency range 0.5 to 4 kHz, increasing to about 4 dB for most of the frequency range above that. The sole exception is at 10 kHz, where the standard error of estimate approaches 6 dB.

Pulsed versus Continuous
Tinnitus Loudness Matches

Choice of Metric

Although it is appropriate to evaluate loudness matches in dB SPL, that technique may increase the overall variance in the data as the resulting measures may confound individual threshold differences with differences in regard

Table 1 Means and Standard Deviations for Hearing Threshold Measurements in dB SPL (N = 26)

<table>
<thead>
<tr>
<th>Stimulation Mode</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulsed Mean</td>
<td>23.2</td>
<td>21.5</td>
<td>21.0</td>
<td>26.3</td>
<td>31.4</td>
<td>53.2</td>
<td>61.6</td>
<td>68.8</td>
<td>65.1</td>
<td>71.2</td>
<td>67.2</td>
</tr>
<tr>
<td>SD</td>
<td>(11.7)</td>
<td>(14.0)</td>
<td>(15.0)</td>
<td>(17.4)</td>
<td>(20.1)</td>
<td>(20.9)</td>
<td>(21.6)</td>
<td>(26.4)</td>
<td>(27.0)</td>
<td>(27.6)</td>
<td>(27.5)</td>
</tr>
<tr>
<td>Continuous Mean</td>
<td>23.3</td>
<td>20.8</td>
<td>21.0</td>
<td>25.0</td>
<td>30.6</td>
<td>53.3</td>
<td>62.5</td>
<td>70.1</td>
<td>65.3</td>
<td>70.5</td>
<td>66.6</td>
</tr>
<tr>
<td>SD</td>
<td>(12.2)</td>
<td>(15.1)</td>
<td>(15.2)</td>
<td>(16.2)</td>
<td>(19.0)</td>
<td>(20.1)</td>
<td>(21.7)</td>
<td>(25.0)</td>
<td>(26.8)</td>
<td>(26.4)</td>
<td>(26.4)</td>
</tr>
<tr>
<td>Difference (pulsed-continuous)</td>
<td>-0.1</td>
<td>0.7</td>
<td>0.0</td>
<td>1.3</td>
<td>0.8</td>
<td>-0.1</td>
<td>-0.9</td>
<td>-1.3</td>
<td>-0.2</td>
<td>0.7</td>
<td>0.6</td>
</tr>
</tbody>
</table>

*Data from 23 subjects, †data from 22 subjects.
The missing data are due to some patients not responding at the highest output of the audiometer at 9 and 10 kHz.
to the loudness of tinnitus. Tinnitus patients often have substantial hearing loss, a tendency that was clearly evident in the present sample of subjects (see Table 1). For that reason, we have chosen to follow the long-standing tradition of evaluating tinnitus loudness matches in dB SL (Fowler, 1940; Reed, 1960; Vernon, 1976), thus normalizing subjects' loudness matches relative to their thresholds. For completeness, however, we also performed all of the analyses described below in terms of dB SPL. The outcomes of the latter analyses were essentially the same as those using dB SL and have therefore been omitted, unless otherwise noted below.
Figure 2  Representative correlation plots for pulsed vs continuous thresholds. Correlations at frequencies 0.5, 2, 4, and 8 kHz are shown. Equations for regression lines (solid lines) are shown in the lower right-hand corner of each plot.

**Mean Loudness Matches**

Figure 3 shows the mean loudness matches, in dB SL, at each frequency for both pulsed and continuous stimuli. The curves almost overlap at most frequencies, emphasizing the similarity of measurements obtained with the two stimulation modes.

Table 3 presents the means and standard deviations for the loudness matches in dB SL. The ANOVA (2 x 9, stimulus mode x frequency, with repeated measures on both factors) did not reveal a significant difference either for the stimulus mode (F[1, 25] = 1.261, p = .272) or the interaction between stimulus mode and frequency (F[8, 200] = 1.307, p = .242).

The relatively small sample sizes for these analyses raises the question whether the samples were large enough to evaluate the statistical significance of small mean differences. Although mean differences in the range 0 to 2 dB are probably not clinically significant, differences larger than about 3 dB might be. There were no a priori data concerning the expected variance of pulsed loudness matches; however, we performed post hoc power analysis using the

---

**Table 2  Relation between Pulsed and Continuous Threshold Measures at Each Frequency: Correlations and Standard Errors of Estimate**

<table>
<thead>
<tr>
<th>Test Frequency (kHz)</th>
<th>Number of Observations (N)</th>
<th>Pearson's r, Pulsed vs Continuous Loudness Matches*</th>
<th>Standard Error of Estimate$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.500</td>
<td>26</td>
<td>.97</td>
<td>2.99</td>
</tr>
<tr>
<td>.750</td>
<td>26</td>
<td>.98</td>
<td>2.91</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>.99</td>
<td>2.21</td>
</tr>
<tr>
<td>1.5</td>
<td>26</td>
<td>.98</td>
<td>3.62</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>.99</td>
<td>2.96</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>.99</td>
<td>3.08</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>.99</td>
<td>3.17</td>
</tr>
<tr>
<td>6</td>
<td>26</td>
<td>.99</td>
<td>3.88</td>
</tr>
<tr>
<td>8</td>
<td>26</td>
<td>.99</td>
<td>3.98</td>
</tr>
<tr>
<td>9</td>
<td>22</td>
<td>.99</td>
<td>4.09</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>.98</td>
<td>5.77</td>
</tr>
</tbody>
</table>

$^*p < .001.$

$S_{pul} = (S_p \sqrt{1-r^2})/\sqrt{N(N-2)}$, where $S_{pul}$ is the standard error of estimate for pulsed thresholds at a given frequency, using continuous thresholds as predictors; $S_p$ is the standard deviation of pulsed thresholds at that frequency; $r$ is the product-moment correlation coefficient at the given frequency, and the last term, $\sqrt{N(N-2)}$, is the correction for small sample size (Downie and Heath, 1965).
Table 3 Means and Standard Deviations for Loudness Match Measurements in dB SL (N = 26)

<table>
<thead>
<tr>
<th>Frequency (kHz)</th>
<th>Pulsed Mean</th>
<th>SD</th>
<th>Continuous Mean</th>
<th>SD</th>
<th>Difference (pulsed-continuous)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>23.5</td>
<td>16.5</td>
<td>20.7</td>
<td>14.3</td>
<td>2.8</td>
</tr>
<tr>
<td>0.75</td>
<td>21.5</td>
<td>15.2</td>
<td>20.2</td>
<td>14.0</td>
<td>1.3</td>
</tr>
<tr>
<td>1</td>
<td>21.6</td>
<td>16.2</td>
<td>21.0</td>
<td>15.5</td>
<td>0.6</td>
</tr>
<tr>
<td>1.5</td>
<td>20.3</td>
<td>16.3</td>
<td>19.9</td>
<td>16.2</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>19.9</td>
<td>17.4</td>
<td>19.5</td>
<td>17.4</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td>12.1</td>
<td>11.2</td>
<td>11.8</td>
<td>9.9</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>11.3</td>
<td>11.8</td>
<td>10.2</td>
<td>11.2</td>
<td>0.1</td>
</tr>
<tr>
<td>6</td>
<td>10.5</td>
<td>11.6</td>
<td>10.5</td>
<td>10.4</td>
<td>0.0</td>
</tr>
<tr>
<td>8*</td>
<td>10.4</td>
<td>11.3</td>
<td>9.8</td>
<td>10.3</td>
<td>0.6</td>
</tr>
<tr>
<td>9'</td>
<td>8.6</td>
<td>13.3</td>
<td>8.4</td>
<td>9.9</td>
<td>1.9</td>
</tr>
<tr>
<td>10'</td>
<td>9.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Data from 24 subjects; †data from 21 subjects.
The missing data are due to some patients not responding at the highest output of the audiometer at 8–10 kHz.

observed variance as an estimate of the population variance. These analyses revealed that, for \( p \leq 0.05 \) (two-tailed), loudness match differences as small as 3 dB were detectable with a power of 0.90.

Although the difference between pulsed and continuous loudness matches was not statistically significant, there was a systematic tendency for the pulsed matches to be larger than the continuous matches (see Fig. 3 and Table 3). Therefore, a Wilcoxon Sign Test for Matched Samples (Welkowitz et al, 1976) was done to determine if this trend was significant. Using this distribution-free test, a significantly greater proportion of higher loudness matches did occur using pulsed stimuli (\( p < 0.01 \)).

The 2 \( \times \) 9 ANOVA did not determine if significant differences occurred between stimulus types at each test frequency. Therefore, orthogonal t-tests were performed for the pairs of mean loudness matches at each frequency. Only the mean loudness matches at 500 Hz, where there was a difference of 2.8 dB, reached significance at the .05 probability level (\( t = 2.75, \) df = 25, \( p < 0.05 \)). However, correcting for multiple t-tests, this difference did not reach a significant level. Nevertheless, at each frequency, the difference between the mean of the pulsed versus the mean of the continuous loudness matches was less than 1 dB for the majority of frequencies. If all 11 of these differences between the pulsed versus continuous means are averaged together, we obtain an overall mean difference of only 0.83 dB SL (0.85 dB for the data expressed in dB SPL).

Individual Data

For each of the test frequencies, each subject's continuous loudness match was subtracted from their pulsed loudness match at that frequency. Figure 4 displays the difference scores

Table 4 Relation between Pulsed and Continuous Loudness Matches at Each Frequency: Correlations and Standard Errors of Estimate

<table>
<thead>
<tr>
<th>Test Frequency (kHz)</th>
<th>Number of Observations (N)</th>
<th>Pearson's r, Pulsed vs Continuous Loudness Matches</th>
<th>Standard Error of Estimate*</th>
</tr>
</thead>
<tbody>
<tr>
<td>.500</td>
<td>26</td>
<td>.92</td>
<td>6.73</td>
</tr>
<tr>
<td>.750</td>
<td>26</td>
<td>.93</td>
<td>5.81</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>.94</td>
<td>5.75</td>
</tr>
<tr>
<td>1.5</td>
<td>26</td>
<td>.93</td>
<td>6.23</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>.95</td>
<td>5.65</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>.94</td>
<td>3.98</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>.96</td>
<td>3.44</td>
</tr>
<tr>
<td>6</td>
<td>26</td>
<td>.89</td>
<td>4.98</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>.94</td>
<td>4.13</td>
</tr>
<tr>
<td>9</td>
<td>21</td>
<td>.93</td>
<td>4.37</td>
</tr>
<tr>
<td>10</td>
<td>21</td>
<td>.93</td>
<td>5.14</td>
</tr>
</tbody>
</table>

*See Table 2 for formula.
Figure 4  Frequency histograms for loudness match differences between the two test modes (pulsed minus continuous). Vertical lines indicate range of values between ±5 dB and ±10 dB.

for the entire group of subjects in the form of 11 histograms stacked vertically (one histogram for each of the 11 test frequencies). In general, most of these individual difference scores were within the range ±5 dB. As with the hearing thresholds, a small number of subjects had difference scores falling outside that range, with a few showing differences larger than ±10 dB.

Pearson's product-moment correlations were computed for the pairs of values represented by each subject's pulsed versus continuous loudness matches at a given frequency. Table 4 summarizes the correlation coefficients obtained at each of the 11 test frequencies, which ranged from 0.89 to 0.96 (p < .001 for all frequencies). Representative scatter plots are depicted in Figure 5 for frequencies 0.5, 2, 4, and 8 kHz.

Table 4 also shows the standard errors of estimate for pulsed loudness matches (with continuous loudness matches as the predictors), computed across the entire frequency range. These values are larger than those for the pulsed thresholds, ranging from 3.44 to 6.73 dB, depending on frequency. In the low-frequency range, these observations are consistent with the fact that many subjects commented on the greater difficulty of the loudness matching task at the lowest frequencies, where the external tone sounded very different from their tinnitus. We have no explanation for the increased standard error of estimate at 10 kHz.
Moreover, some pathological ears demonstrate more rapid auditory adaptation (Harbert et al., 1968), and, in those cases, pulsed-tone thresholds might be expected to show even greater sensitivity (Wright, 1969; Yantis, 1985). Some investigations comparing auditory thresholds obtained with pulsed versus continuous stimuli have reported roughly equivalent thresholds in normal-hearing ears, but significantly improved thresholds using pulsed tones in subjects with cochlear and retrocochlear hearing loss (Jerger, 1960; Hood, 1972). That phenomenon, however, was not observed in the patients who participated in the present study, most of whom had substantial hearing impairment in the test ear. On the contrary, many of the subjects for the present study exhibited better threshold sensitivity for continuous than for pulsed tones (note the values to the left of 0 dB in Fig. 1).

Several investigators have suggested that pulsed tones should be used when evaluating pure-tone hearing thresholds in tinnitus patients, to help the patient distinguish between the externally presented tones and the internally generated tinnitus sensation (Douek and Reid, 1968; Green, 1972b; Fulton and Lloyd, 1975; Yantis, 1994). The implication here is that thresholds should be higher (worse) with continuous tones than with pulsed tones. Two previous investigations have been conducted to test that premise. Hochberg and Waltzman (1972) measured pulsed and continuous thresholds in 25 patients with sensorineural hearing loss and tinnitus and in a control group of 25 normal-hearing individuals who did not have tinnitus. The differences between mean thresholds for pulsed and continuous conditions did not exceed 1.6 dB at any of the test frequencies (0.125–8 kHz).

Mineau and Schlauch (1997) evaluated hearing thresholds in one group of patients with tinnitus using pulsed tones and another group using continuous tones. Because the two sets of measurements were not obtained in the same subjects, but instead were obtained in two different groups, this introduced a potential group difference that would confound the comparison. In any event, neither of these studies showed that thresholds obtained with pulsed tones were lower than those obtained with continuous tones for patients with tinnitus. Mineau and Schlauch (1997), however, noted that significantly more false-positive responses occurred with continuous tones than with pulsed tones. They also found that more stimulus presentations were necessary using continuous tones than using pulsed tones, which was significant only at one
test frequency (4 kHz). Finally, Hochberg and Waltzman (1972) noted that most of their patients with tinnitus reported a preference for the pulsed tones. However, almost the same proportion of subjects from their control group (without tinnitus) indicated the same preference.

In agreement with results from the studies of Hochberg and Waltzman (1972) and Mineau and Schlauch (1997), the present results tend to contradict the suggestion that pulsed tones will result in improved thresholds in patients with tinnitus. Although the duration of the test stimuli had relatively little effect on threshold measurements in most tinnitus patients, if anything the continuous thresholds appeared more sensitive. For the small number of the present study’s subjects in whom pulsed thresholds differed from continuous thresholds by more than 5 to 10 dB, the direction of the differences was about equally divided between positive and negative values, thus indicating little or no systematic trend in favor of either pulsed or continuous stimuli.

Pulsed versus Continuous Loudness Matches

The present results provide a comprehensive evaluation of tinnitus loudness matches over a wide range of test frequencies and corroborate earlier reports (Goodwin and Johnson, 1980; Tyler and Conrad-Armes, 1983; Vernon and Meikle, 1988; Risey et al, 1989; Meikle et al, 1995) indicating that the size of the loudness match is frequently smaller at the tinnitus pitch than at lower frequencies with more normal hearing (see Fig. 3). The present results also extend loudness match observations into the frequency range above each subject’s “tinnitus frequency,” finding that the mean loudness matches tended to remain relatively constant for frequencies at and above 3 kHz.

Although ANOVA did not reveal significant differences between the mean loudness matches for pulsed versus continuous stimuli, there was a systematic trend in that eight of the nine mean differences were positive (see Table 3). The results of the Wilcoxon test, which confirmed the statistical significance of this trend, support the conclusion that pulsed loudness matches tend to be slightly greater than those obtained with continuous tones. The differences are small enough, however, that a larger sample of subjects would appear to be necessary in order to demonstrate statistical significance using ANOVA.

When viewed from a different perspective, the present results could actually be used as a demonstration that many subjects show good test–retest reliability when asked to provide loudness matches for their tinnitus, regardless of the stimulus duration. Despite the intertest interval of 10 to 15 minutes, the data shown in Figure 4 reveal that many subjects were capable of reproducing their loudness matches to within ±2 to 3 dB (Bailey, 1979; Vernon et al, 1980). It is not clear why other studies have reported poorer test–retest reliability (Penner, 1983, 1986; Burns, 1984); however, it seems possible that methodological differences could exert a strong influence on the intrasession reliability of these measures. For example, some studies used smaller numbers of subjects and did not exclude subjects whose tinnitus was described as “fluctuating.” The present study employed 26 subjects who were selected on the basis that they described their tinnitus as “non-fluctuating.”

The greater variability at 500 and 750 Hz may be due to greater difficulty comparing the loudness of low-frequency tones to the loudness of high-frequency tinnitus (subjects volunteered that it was difficult). Subjects commented “That doesn’t sound at all like my tinnitus” and “It’s hard to compare them.”

Group versus Individual Data

Figures 1 and 4 both reveal a small number of individuals who show relatively large differences (>10 dB) between their pulsed and continuous measures. These outliers were examined to determine whether there were individual attributes that might be associated with such response discrepancies. Neither tinnitus etiology, tinnitus pitch, or other individual differences appeared to account for the discrepancies.

Closer examination revealed that the large discrepancies were contributed primarily by nine individuals, or approximately one-third of the group. (Six of these nine individuals showed consistently positive or negative values, and three varied widely from positive to negative.) The remaining 17 subjects showed only small variations in the differences between their dB SL loudness matches. Those 17 are the subjects primarily responsible for the finding that there were no significant differences between the two stimulus modes in regard to loudness matches. Data from the remaining nine subjects exhibited no systematic trend, as the number showing positive differences was about equal to the num-
ber showing negative differences. The fact that these larger differences in a few subjects were seen across all frequencies suggests that this was not due to a procedural effect such as earphone placement or choice of test frequency.

The net conclusion to be derived is that studies employing group data may be compared with confidence; contrariwise, it may be hazardous to generalize from studies based on data from a small number of subjects. Studies employing only a few subjects may be vulnerable to disproportionate representation of the values obtained from the outliers. These results reveal the potential importance of stimulus duration for a small subset of the patient population; it remains to be determined what sets the outliers apart from the larger group of patients for whom the stimulus duration makes little difference.

CONCLUSIONS

The present study shows that for the majority of individuals, loudness matches for tinnitus are relatively independent of the duration of the test stimuli, at least for stimulus durations less than several seconds. However, studies that have used small numbers of subjects should be viewed as being more vulnerable to the potential biasing effects of outliers.

The present results may have relevance for the evaluation of various types of relief procedures for tinnitus, specifically for measuring the magnitude of tinnitus before and after treatment. Thus, either pulsed or continuous stimuli can be used with confidence to evaluate effectiveness of tinnitus therapy as long as the group of subjects is relatively large. Again, if the group of subjects is small, the potential disproportionate effect of outliers can contaminate the test results.

As tinnitus becomes an increasing national health concern, due to widespread noise exposure and other environmental factors that contribute to the increased incidence of hearing loss and tinnitus, the ability to obtain reliable measures of the magnitude of tinnitus has far-reaching implications. It is hoped that the present results will contribute to greater awareness of the influence of testing technique upon the obtained measures.

Acknowledgment. Funding for this study was provided by the National Institute of Mental Health (NIMH), U.S. Department of Education, and the Veterans Affairs Rehabilitation Research and Development (RR&D) Service.

REFERENCES


